

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

## **IMAGES ARE BEST AVAILABLE COPY.**

As rescanning documents *will not* correct images,  
Please do not report the images to the  
Image Problem Mailbox.

## PATENT ABSTRACTS OF JAPAN

(11)Publication number : 08-289297

(43)Date of publication of application : 01.11.1996

(51)Int.Cl.

H04N 7/30

H04N 5/92

(21)Application number : 07-089613

(71)Applicant : HITACHI LTD

(22)Date of filing : 14.04.1995

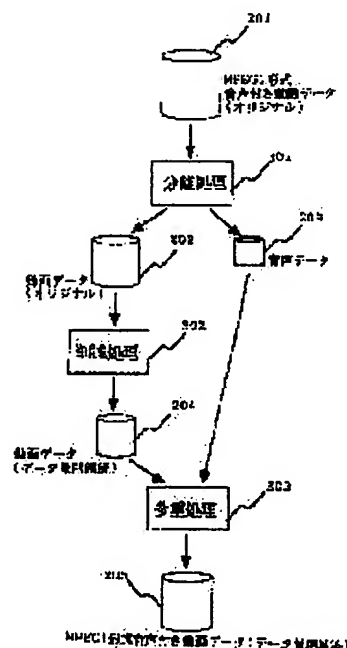
(72)Inventor : NONOMURA ITARU  
 HASHIMOTO SHINICHI  
 IKETANI YUICHI  
 YAMADA TAKEHIRO  
 TANAKA KAZUAKI  
 IMAI YASUHIRO

## (54) ANIMATION DATA PREPARING DEVICE WITH VOICE

## (57)Abstract:

**PURPOSE:** To prepare the animation data with voice corresponding to desired reproduction time which has data quantity less than animation data with voice and is suitable to the simultaneous outputs of the corresponded animation and voice from animation data with voice.

**CONSTITUTION:** A separation processing 301 separates animation data with original voice into animation data and voice data. A deletion processing 302 deletes the part showing the high frequency component of the animation data within each frame composing animation data when the animation data separated by the separation processing 301 is the animation data encoded by an orthogonal conversion system by a frame unit, for instance. A multiple processing 303 multiplexes the part of the animation data in which the frame of the number according to designated reproduction time is encoded (reproduction object animation data) of the animation data that the deletion pressing 302 deletes and the part to be simultaneously reproduced with reproduction object animation data of the voice data separated by the separation processing 301.



## LEGAL STATUS

[Date of request for examination]

16.02.2001

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the  
 examiner's decision of rejection or application  
 converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of  
 rejection]

[Date of requesting appeal against examiner's decision]

of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平8-289297

(43) 公開日 平成8年(1996)11月1日

(51) Int.Cl. <sup>6</sup>	識別記号	庁内整理番号	F I	技術表示箇所
H 0 4 N 7/30 5/92			H 0 4 N 7/133 5/92	Z H

審査請求 未請求 請求項の数 8 O L (全 16 頁)

(21) 出願番号 特願平7-89613

(22) 出願日 平成7年(1995)4月14日

(71) 出願人 000005108  
株式会社日立製作所  
東京都千代田区神田駿河台四丁目6番地

(72) 発明者 野々村 到  
神奈川県川崎市麻生区王禅寺1099番地 株式会社日立製作所システム開発研究所内

(72) 発明者 橋本 真一  
神奈川県川崎市麻生区王禅寺1099番地 株式会社日立製作所システム開発研究所内

(72) 発明者 池谷 勇一  
神奈川県川崎市麻生区王禅寺1099番地 株式会社日立製作所システム開発研究所内

(74) 代理人 弁理士 富田 和子

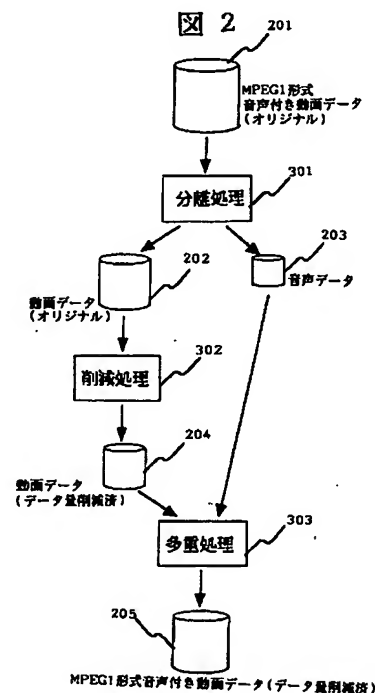
最終頁に続く

(54) 【発明の名称】 音声付き動画データ作成装置

(57) 【要約】

【目的】 音声付き動画データから、該音声付き動画データよりデータ量が少なく、かつ、対応する動画と音声とを同時に出力するのに適した、所望の再生時間分の音声付き動画データを作成する。

【構成】 分離処理301は、オリジナルの音声付き動画データを動画データと音声データとに分離し、削減処理302は、例えば、分離処理301が分離した動画データが、フレーム単位に直交変換方式で符号化された動画データである場合には、該動画データを構成する各フレーム内の動画データのうちの高周波数成分を表す部分を削除する。多重処理303は、削減処理302が削減した動画データのうちの、指定された再生時間に応じた個数のフレームを符号化した動画データの部分（再生対象動画データ）と、分離処理301が分離した音声データのうちの、再生対象動画データと同時に再生されるべき部分とを多重する。



## 1

## 【特許請求の範囲】

【請求項 1】複数のフレームからなる動画データをフレーム単位に直交変換方式で符号化した動画データと、音声データとが多重された音声付き動画データを、動画データと音声データとに分離する分離手段と、

上記分離手段が分離した動画データについて、各フレーム内のデータのうちの高周波数成分を表すデータを削除することにより、各フレーム内のデータ量を削減する削減手段と、

上記削減手段が削減した動画データのうちの、指定された再生時間に応じた個数のフレームを符号化した動画データの部分である再生対象動画データと、上記分離手段が分離した音声データのうちの、上記再生対象動画データと同時に再生されるべき部分である再生対象音声データとを多重する多重手段とを有することを特徴とする音声付き動画データ作成装置。

【請求項 2】請求項 1 記載の音声付き動画データ作成装置において、

上記多重手段は、

上記削減手段が削減した動画データについて、フレームごとに、該フレーム内の先頭の動画データの再生開始時刻情報を含む第 1 の補助データを作成する第 1 の作成手段と、

上記分離手段が分離した音声データについて、同時に再生されるべきフレームごとに、該フレーム内の先頭の音声データの再生開始時刻情報を含む第 2 の補助データを作成する第 2 の作成手段と、

上記第 1 の作成手段が作成した第 1 の補助データに基づいて、上記削減手段が削減した動画データのうちから、指定された再生時間に応じた個数のフレームを抽出することにより、上記再生対象動画データを決定する第 1 の抽出手段と、

上記第 1 の作成手段が作成した第 1 の補助データおよび上記第 2 の作成手段が作成した第 2 の補助データに基づいて、上記分離手段が分離した音声データのうちから、上記第 1 の抽出手段が決定した再生対象動画データと再生開始時刻が対応するフレームを抽出することにより、上記再生対象音声データを決定する第 2 の抽出手段とを有し、

上記第 1 の抽出手段が抽出した再生対象動画データと上記第 2 の抽出手段が抽出した再生対象音声データとを多重することを特徴とする音声付き動画データ作成装置。

【請求項 3】複数のフレームからなる動画データをフレーム単位にフレーム間予測符号化方式で符号化し、フレーム内の動画データを他のフレーム内の動画データとは独立に符号化したフレームである 1 個の I ピクチャ、前方向にある 1 個の I ピクチャから順方向予測符号化したフレームである 1 個以上の P ピクチャ、前方向にある 1 個の I ピクチャまたは P ピクチャと後方向にある 1 個の I ピクチャまたは P ピクチャとから双方向予測符号化し

## 2

たフレームである 1 個以上の B ピクチャからなる再生処理単位である GOP (Group Of Picture) を複数含む動画データと、音声データとが多重された音声付き動画データを、動画データと音声データとに分離する分離手段と、

上記分離手段が分離した動画データについて、上記 B ピクチャ内のデータを、所定の値のデータに置換することにより、上記 B ピクチャ内のデータ量を削減する削減手段と、

10 上記削減手段が削減した動画データについて、指定された再生時間に応じた個数の GOP を抽出することにより、再生すべき動画データである再生対象動画データを決定すると共に、上記再生対象動画データと、上記分離手段が分離した音声データのうちの、上記再生対象動画データと同時に再生されるべき部分である再生対象音声データとを多重する多重手段とを有することを特徴とする音声付き動画データ作成装置。

【請求項 4】請求項 3 記載の音声付き動画データ作成装置において、

20 上記多重手段は、

上記削減手段が削減した動画データについて、上記 GOP ごとに、該 GOP 内の先頭の動画データの再生開始時刻情報を含む第 1 の補助データを作成する第 1 の作成手段と、

上記削減手段が削減した動画データについて、上記ピクチャごとに、該ピクチャ内の先頭の動画データの再生開始時刻情報を含む第 2 の補助データを作成する第 2 の作成手段と、

30 上記分離手段が分離した音声データについて、同時に再生されるべきフレームごとに、該フレーム内の先頭の音声データの再生開始時刻情報を含む第 3 の補助データを作成する第 3 の作成手段と、

上記第 1 の作成手段が作成した第 1 の補助データに基づいて、上記削減手段が削減した動画データのうちから、指定された再生時間に応じた個数の GOP を抽出することにより、上記再生対象動画データを決定する第 1 の抽出手段と、

40 上記第 2 の作成手段が作成した第 2 補助データおよび上記第 3 の作成手段が作成した第 3 の補助データに基づいて、上記分離手段が分離した音声データのうちから、上記第 1 の抽出手段が決定した再生対象動画データと再生開始時刻が対応するフレームを抽出することにより、上記再生対象音声データを決定する第 2 の抽出手段とを有し、

上記第 1 の抽出手段が抽出した再生対象動画データと上記第 2 の抽出手段が抽出した再生対象音声データとを多重することを特徴とする音声付き動画データ作成装置。

【請求項 5】請求項 3 記載の音声付き動画データ作成装置において、

50 上記多重手段は、上記 GOP の切れ目が、音声付き動画

## 3

データを伝送する際の伝送処理単位の切れ目に一致するようにして多重することを特徴とする音声付き動画データ作成装置。

【請求項 6】請求項 4 または 5 記載の音声付き動画データ作成装置において、

上記第 1 の作成手段は、上記削減手段が削減した動画データのフレームレートと、当該 GOP より前にある全ての GOP に含まれているピクチャの数とに基づいて、上記 GOP ごとの再生開始時刻情報を算出し、

上記第 2 の作成手段は、上記削減手段が削減した動画データのフレームレートと、上記 GOP ごとの再生開始時刻情報と、当該ピクチャを含む GOP における該ピクチャの再生順序とに基づいて、上記ピクチャごとの再生開始時刻情報を算出し、

上記第 3 の作成手段は、上記分離手段が分離した音声データのサンプリング周波数と、当該フレームより前にあるフレームに含まれているサンプル数とに基づいて、フレームごとの再生開始時刻情報を算出することを特徴とする音声付き動画データ作成装置。

【請求項 7】複数のフレームからなる動画データをフレーム単位に直交変換方式で符号化した動画データと、音声データとが多重された音声付き動画データを、動画データと音声データとに分離する分離手段と、

上記分離手段が分離した動画データについて、各フレーム内のデータのうちの高周波数成分を表すデータを削除することにより、各フレーム内のデータ量を削減する削減手段と、

上記削減手段が削減した動画データのうちの、指定された再生時間に応じた個数のフレームを符号化した動画データの部分である再生対象動画データと、上記分離手段が分離した音声データのうちの、上記再生対象動画データと同時に再生されるべき部分である再生対象音声データとを多重する多重手段とを有する音声付き動画データ作成装置と、

上記音声付き動画データ作成装置が作成した音声付き動画データをネットワークを介してクライアントに伝送する伝送装置とを備えたことを特徴とするビデオサーバ。

【請求項 8】複数のフレームからなる動画データをフレーム単位にフレーム間予測符号化方式で符号化し、フレーム内の動画データを他のフレーム内の動画データとは独立に符号化したフレームである 1 個の I ピクチャ、前方向にある 1 個の I ピクチャから順方向予測符号化したフレームである 1 個以上の P ピクチャ、前方向にある 1 個の I ピクチャまたは P ピクチャと後方向にある 1 個の I ピクチャまたは P ピクチャとから双方向予測符号化したフレームである 1 個以上の B ピクチャからなる再生処理単位である GOP (Group Of Picture) を複数含む動画データと、音声データとが多重された音声付き動画データを、動画データと音声データとに分離する分離手段と、

## 4

上記分離手段が分離した動画データについて、上記 B ピクチャ内のデータを、所定の値のデータに置換することにより、上記 B ピクチャ内のデータ量を削減する削減手段と、

上記削減手段が削減した動画データについて、指定された再生時間に応じた個数の GOP を抽出することにより、再生すべき動画データである再生対象動画データを決定すると共に、上記再生対象動画データと、上記分離手段が分離した音声データのうちの、上記再生対象動画データと同時に再生されるべき部分である再生対象音声データとを多重する多重手段とを有する音声付き動画データ作成装置と、

上記音声付き動画データ作成装置が作成した音声付き動画データをネットワークを介してクライアントに伝送する伝送装置とを備えたことを特徴とするビデオサーバ。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は、音声データと動画データとが多重された音声付き動画データから、該音声付き動画データよりデータ量が少なく、かつ、対応する音声と動画とを同時に出力するのに適した、所望の再生時間分の音声付き動画データを作成することが可能な音声付き動画データ作成装置に関する。

【0002】

【従来の技術】近年、オフィスにおける LAN (Local Area Network) 等のネットワークインフラの普及および高性能化や、インターネットへのアクセス環境の整備に伴い、複数の情報処理装置間で音声付き動画データをネットワークを介して伝送するマルチメディア伝送システムが普及する兆しがある。

【0003】マルチメディア伝送システムにおいては、音声付き動画データの供給源である情報処理装置（以下、ビデオサーバと称す。）が、動画データと音声データとを多重して作成した音声付き動画データを蓄積し、蓄積した音声付き動画データをネットワークを介して他の情報処理装置（以下、クライアントと称す。）に伝送するようになっている。なお、クライアントは、ビデオサーバからネットワークを介して伝送された音声付き動画データを再生するが、この際に、対応する動画と音声とが同時に出力されることが好ましい。

【0004】また、パーソナルコンピュータやワークステーションの性能の向上に伴い、パーソナルコンピュータやワークステーションをクライアントとして用い、パーソナルコンピュータやワークステーションで、手軽に音声付き動画データを再生することができるようになっている。

【0005】ところで、動画データや音声データは、データ量が非常に大きいことから、その蓄積や伝送を行う際には、圧縮処理を施すことにより、蓄積に必要な記憶装置の記憶容量を削減したり、伝送に必要な時間を短縮

することが、一般に行われている。

【0006】動画データや音声データを圧縮する圧縮方式は多数あるが、そのうちの代表的なものとしては、例えば、ISOによって勧告された、動画データを圧縮するMPEG1 (Moving Picture experts Group Phase 1) ビデオ符号化規格や音声データを圧縮するMPEG1 オーディオ符号化規格がある。また、MPEG1 規格に準拠した動画データと音声データとの多重方法を定めた規格として、MPEG1 システム多重化規格がある。

【0007】これらの規格の仕様や圧縮方式およびデータ構造等の概要については、「ポイント図解最新MPEG教科書」, 1994年8月1日, アスキー社刊のP89~P128およびP231~P253に記載されている。

【0008】しかしながら、処理能力が低い情報処理装置をクライアントとして用いたり低速なネットワークを用いているマルチメディア伝送システムにおいては、圧縮された音声付き動画データであっても、依然として、伝送や復号等の処理に時間がかかり、音声付き動画データをリアルタイムで再生することができないという問題点がある。

【0009】例えば、ビデオサーバが、毎秒1.5Mビットのデータ量を持つMPEG1 準拠の音声付き動画データを、毎秒64Kビットの伝送速度のN-ISDN (Narrow-Integrated Service Digital Network) 回線で、クライアントに伝送する場合は、再生時間の約24倍の伝送時間が必要となるので、クライアントにおいて、音声付き動画データをリアルタイムで再生することは不可能である。

【0010】そこで、クライアントにおいて、ビデオサーバからネットワークを介して伝送された音声付き動画データを、一旦記憶装置に記憶してから、後で再生するようにしようとする、クライアントに大容量の記憶装置を設ける必要がある。

【0011】例えば、クライアントにおいて、毎秒1.5Mビットのデータ量を持つMPEG1 準拠の音声付き動画データを1時間分だけ蓄積するためには、675Mバイトの記憶容量が必要となる。

【0012】一方、音声付き動画データを再生する際に、内容の把握を短時間でを行うために、音声付き動画データを早送り再生したいというニーズがある。

【0013】音声付き動画データの早送り再生を実現するためには、復号等の処理の高速化を図ればよいが、処理負荷の面で困難であるので、音声付き動画データのうちの動画データについて、その一部を再生して残りをスキップするという処理の繰り返すことによって、動画データの早送り再生を実現することが一般的である。

【0014】ところが、動画データは、その途中から再生する場合に、動画データのフレーム (画面) の切れ目といった、意味のある切れ目から再生を開始しないと、

ノイズが発生して正常に再生されない。従って、スキップ終了後に動画データを正常に再生するためには、動画データの意味のある切れ目を検出し、検出した切れ目から再生を開始する必要がある。早送り再生は、通常再生よりも処理負荷が高いという問題点がある。

【0015】上述した問題点を解決するための従来技術として、動画データのデータ量を削減する技術、および、早送り再生専用の動画データを作成する技術が開示されている。

【0016】例えば、特開平6-70174号公報においては、動画データ中の高周波成分を削除することにより、動画データのデータ量を削減する技術が開示されている。

【0017】また、例えば、特開平6-133263号公報においては、事前にオリジナルの動画データを解析して早送り再生専用の動画データを作成しておき、早送り再生時には、早送り再生専用の動画データを再生することにより、処理負荷を軽減する技術が開示されている。

【0018】

【発明が解決しようとする課題】再生時の音声の有無は、内容の把握のしやすさに大きな影響があるので、早送り再生であっても、対応する音声と動画が同時に出力されることが望ましい。

【0019】ところが、上述した従来技術のうち、特開平6-70174号公報に記載されている技術においては、動画データ中の高周波成分を削除することにより、動画データのデータ量を削減する技術が開示されているが、音声データに関する考慮はなされていない。

【0020】また、特開平6-133262号公報に記載されている技術においては、事前にオリジナルの動画データを解析して早送り再生専用の動画データを作成しておき、早送り再生時には、早送り再生専用データを再生することにより、処理負荷を軽減する技術が開示されているが、音声データに関する考慮はなされていない。

【0021】音声データも、動画データと同様に、その途中から再生する場合に、音声データの復号処理単位といった、意味のある切れ目から再生を開始しないと、ノイズが発生して正常に再生されない。従って、スキップ終了後に音声データを正常に再生するためには、音声データの意味のある切れ目を検出し、検出した切れ目から再生を開始する必要がある。

【0022】そこで、本発明の目的は、音声データと動画データとが多重された音声付き動画データから、該音声付き動画データよりデータ量が少なく、かつ、対応する音声と動画とを同時に出力するのに適した、所望の再生時間分の音声付き動画データを作成することが可能な音声付き動画データ作成装置を提供することにある。

【0023】

【課題を解決するための手段】上記目的を達成するため

10

20

30

40

50

## 7

に、本発明は、(1) 複数のフレームからなる動画データをフレーム単位に直交変換方式で符号化した動画データと、音声データとが多重された音声付き動画データを、動画データと音声データとに分離する分離手段、

(2) 上記分離手段が分離した動画データについて、各フレーム内のデータのうちの高周波数成分を表すデータを削除することにより、各フレーム内のデータ量を削減する削減手段、(3) 上記削減手段が削減した動画データのうちの、指定された再生時間に応じた個数のフレームを符号化した動画データの部分である再生対象動画データと、上記分離手段が分離した音声データのうちの、上記再生対象動画データと同時に再生されるべき部分である再生対象音声データとを多重する多重手段、を有する音声付き動画データ作成装置を提供している。

【0024】上記多重手段は、詳しくは、(1) 上記削減手段が削減した動画データについて、フレームごとに、該フレーム内の先頭の動画データの再生開始時刻情報を含む第1の補助データを作成する第1の作成手段、

(2) 上記分離手段が分離した音声データについて、同時に再生されるべきフレームごとに、該フレーム内の先頭の音声データの再生開始時刻情報を含む第2の補助データを作成する第2の作成手段、(3) 上記第1の作成手段が作成した第1の補助データに基づいて、上記削減手段が削減した動画データのうちから、指定された再生時間に応じた個数のフレームを抽出することにより、上記再生対象動画データを決定する第1の抽出手段、

(4) 上記第1の作成手段が作成した第1の補助データおよび上記第2の作成手段が作成した第2の補助データに基づいて、上記分離手段が分離した音声データののうちから、上記第1の抽出手段が決定した再生対象動画データと再生開始時刻が対応するフレームを抽出することにより、上記再生対象音声データを決定する第2の抽出手段、を有し、上記第1の抽出手段が抽出した再生対象動画データと上記第2の抽出手段が抽出した再生対象音声データとを多重するようにしている。

【0025】また、本発明は、(1) 複数のフレームからなる動画データをフレーム単位にフレーム間予測符号化方式で符号化し、フレーム内の動画データを他のフレーム内の動画データとは独立に符号化したフレームである1個のIピクチャ、前方向にある1個のIピクチャから順方向予測符号化したフレームである1個以上のPピクチャ、前方向にある1個のIピクチャまたはPピクチャと後方向にある1個のIピクチャまたはPピクチャとから双方向予測符号化したフレームである1個以上のBピクチャからなる再生処理単位であるGOP (Group Of Picture) を複数含む動画データと、音声データとが多重された音声付き動画データを、動画データと音声データとに分離する分離手段、(2) 上記分離手段が分離した動画データについて、上記Bピクチャ内のデータを、所定の値のデータに置換することにより、上記Bピクチャ

## 8

ャ内のデータ量を削減する削減手段、(3) 上記削減手段が削減した動画データについて、指定された再生時間に応じた個数のGOPを抽出することにより、再生すべき動画データである再生対象動画データを決定すると共に、上記再生対象動画データと、上記分離手段が分離した音声データのうちの、上記再生対象動画データと同時に再生されるべき部分である再生対象音声データとを多重する多重手段、を有する音声付き動画データ作成装置を提供している。

【0026】上記多重手段は、詳しくは、(1) 上記削減手段が削減した動画データについて、上記GOPごとに、該GOP内の先頭の動画データの再生開始時刻情報を含む第1の補助データを作成する第1の作成手段、

(2) 上記削減手段が削減した動画データについて、上記ピクチャごとに、該ピクチャ内の先頭の動画データの再生開始時刻情報を含む第2の補助データを作成する第2の作成手段、(3) 上記分離手段が分離した音声データについて、同時に再生されるべきフレームごとに、該フレーム内の先頭の音声データの再生開始時刻情報を含む第3の補助データを作成する第3の作成手段、(4) 上記第1の作成手段が作成した第1の補助データに基づいて、上記削減手段が削減した動画データののうちから、指定された再生時間に応じた個数のGOPを抽出することにより、上記再生対象動画データを決定する第1の抽出手段、(5) 上記第2の作成手段が作成した第2補助データおよび上記第3の作成手段が作成した第3の補助データに基づいて、上記分離手段が分離した音声データののうちから、上記第1の抽出手段が決定した再生対象動画データと再生開始時刻が対応するフレームを抽出することにより、上記再生対象音声データを決定する第2の抽出手段、を有し、上記第1の抽出手段が抽出した再生対象動画データと上記第2の抽出手段が抽出した再生対象音声データとを多重するようにしている。

【0027】なお、上記多重手段は、上記GOPの切れ目が、音声付き動画データを伝送する際の伝送処理単位の切れ目に一致するようにして多重することが好ましい。

【0028】また、上記第1の作成手段は、上記削減手段が削減した動画データのフレームレートと、当該GOPより前にある全てのGOPに含まれているピクチャの数とに基づいて、上記GOPごとの再生開始時刻情報を算出することができ、上記第2の作成手段は、上記削減手段が削減した動画データのフレームレートと、上記GOPごとの再生開始時刻情報と、当該ピクチャを含むGOPにおける該ピクチャの再生順序とに基づいて、上記ピクチャごとの再生開始時刻情報を算出することができ、上記第3の作成手段は、上記分離手段が分離した音声データのサンプリング周波数と、当該フレームより前にあるフレームの数とに基づいて、フレームごとの再生開始時刻情報を算出することができる。



【0029】

【作用】本発明の音声付き動画データ作成装置においては、上記削減手段は、上記分離手段が分離した動画データについて、各フレーム内のデータのうちの高周波数成分を表すデータを削除することにより、各フレーム内のデータ量を削減する。

【0030】そして、上記多重手段は、上記削減手段が削減した動画データのうちの、指定された再生時間に応じた個数のフレームを符号化した動画データの部分である再生対象動画データと、上記分離手段が分離した音声データのうちの、上記再生対象動画データと同時に再生されるべき部分である再生対象音声データとを多重する。

【0031】また、上記削減手段は、上記分離手段が分離した動画データについて、上記Bピクチャ内のデータを、所定の値のデータに置換することにより、上記Bピクチャ内のデータ量を削減する。

【0032】そして、上記削減手段が削減した動画データについて、指定された再生時間に応じた個数のGOPを抽出することにより、再生すべき動画データである再生対象動画データを決定すると共に、上記再生対象動画データと、上記分離手段が分離した音声データのうちの、上記再生対象動画データと同時に再生されるべき部分である再生対象音声データとを多重する。

【0033】従って、本発明の音声付き動画データ作成装置によれば、音声付き動画データから、該音声付き動画データよりデータ量が少なく、かつ、対応する動画と音声と同時に出力するのに適した、所望の再生時間分の音声付き動画データを作成することができる。

【0034】

【実施例】以下、本発明の実施例について図面を参照して説明する。

【0035】まず、本発明の第1の実施例について説明する。

【0036】なお、以下の説明では、音声付き動画データがMPEG1形式の音声付き動画データである場合を例にとって説明する。

【0037】図1は第1の実施例を適用したビデオサーバの構成を示すブロック図である。

【0038】図1に示すように、ビデオサーバは、CPU101と、メインメモリ102と、記憶装置103と、伝送装置104と、バス105とを備えて構成されている。

【0039】本実施例において、ビデオサーバは、記憶装置103に格納されているMPEG1形式のオリジナルの音声付き動画データから、該オリジナルの音声付き動画データよりデータ量が少なく、かつ、対応する動画と音声とを同時に出力するのに適した、所望の再生時間分のMPEG1形式の音声付き動画データを作成し、作成したMPEG1形式の音声付き動画データを記憶装置

103に格納する音声付き動画データ作成処理を行う。

【0040】なお、本実施例は、ビデオサーバが行う音声付き動画データ作成処理をソフトウェアで実現した例である。すなわち、音声付き動画データ作成処理は、CPU101が、記憶装置103に格納されているソフトウェアをメインメモリ102にロードし、メインメモリ102上でロードしたソフトウェアを実行することによって実現される。

【0041】なお、メインメモリ102は、半導体メモリ等の記憶媒体を用いて構成された揮発性の記憶装置であり、記憶装置103は、磁気記憶装置等の記憶媒体を用いて構成された不揮発性の記憶装置である。

【0042】また、伝送装置104は、音声付き動画データ作成処理によって作成されて記憶装置103に格納されたMPEG1形式の音声付き動画データを、ネットワークを介してクライアントに伝送する。

【0043】また、各ブロック間でバス105を介して行われるデータの転送は、全てCPU101が制御しているものである。

【0044】さらに、キーボードやマウス等の入力装置、および、CRT等の表示装置を備えるようにしてもよい。

【0045】図3は音声付き動画データ作成処理の処理手順を示すフローチャートである。

【0046】図3に示すように、音声付き動画データ作成処理においては、分離処理301、削減処理302、多重処理303という順序で処理が行われる。

【0047】図2は音声付き動画データ作成処理におけるデータの流れを示す図である。

【0048】図2において、ファイル201は、MPEG1形式のオリジナルの音声付き動画データが格納されたファイルを示している。

【0049】また、ファイル202は、分離処理301によって、ファイル201に格納されているオリジナルの音声付き動画データから分離された動画データが格納されたファイルを示しており、ファイル203は、分離処理301によって、ファイル201に格納されているオリジナルの音声付き動画データから分離された音声データが格納されたファイルを示している。

【0050】また、ファイル204は、削減処理302によって、ファイル202に格納されている動画データのデータ量が削減された動画データが格納されたファイルを示している。

【0051】また、ファイル205は、多重処理303によって、ファイル204に格納されている動画データとファイル203に格納されている音声データとが多重された、MPEG1形式の音声付き動画データが格納されたファイルを示している。ファイル205に格納されているMPEG1形式の音声付き動画データが、ネットワークを介してクライアントに伝送されることとなる。

【0052】なお、これらのファイル201~205は、記憶装置103に格納されるものである。

【0053】以下、音声付き動画データ作成処理の詳細について説明する。

【0054】まず、分離処理301について、図4および図5を用いて説明する。

【0055】図4はMPEG1形式の音声付き動画データの構造図である。

【0056】図4に示すように、MPEG1形式の音声付き動画データは、バック400およびパケット410の2層からなる階層構造となっている。

【0057】バック400は、バックヘッダ401と、システムヘッダ402と、1個以上のパケット410とからなる。

【0058】ここで、バックヘッダ401は、バック400の先頭を示すバック開始コード（同期コードの一種である。）、後述するタイムスタンプに時間基準を与えるためのシステム時刻基準参照値、多重化レート等からなる。また、システムヘッダ402は、システムヘッダ402の先頭を示すシステムヘッダ開始コード（同期コードの一種である。）、ビットレート、動画データのチャンネル数、音声データのチャンネル数等からなる。

【0059】また、パケット410は、パケットヘッダ411と、データ（動画データまたは音声データ）412とからなる。

【0060】ここで、パケットヘッダ411は、パケット310の先頭を示すパケット開始コード（同期コードの一種である。）、対応する音声と動画とを同時に出力するために必要な時刻情報であるタイムスタンプ等からなる。なお、タイムスタンプには、いつ再生すべきかを示す再生の時刻管理情報、および、いつ復号すべきかを示す復号の時刻管理情報の2種類がある。また、パケット開始コードには、データ412のデータ種別が含まれている。

【0061】また、バック400の最後には、音声付き動画データの終了を示す終了コード（同期コードの一種である。）420が付加されている。

【0062】図5は分離処理301の処理手順を示すフローチャートである。

【0063】分離処理301においては、図5に示すように、まず、ファイル201に格納されているオリジナルのMPEG1形式の音声付き動画データから、MPEG1システム多重化規格で定められた同期コードを検出する（ステップ501）。

【0064】続いて、ステップ501で検出した同期コードが終了コード420であるか否かを判定し（ステップ502）、終了コード420である場合は、処理を終了する。

【0065】また、終了コード420でない場合は、ステップ501で検出した同期コードがパケット開始コー

ド（パケットヘッダ411に含まれる。）であるか否かを判定し（ステップ503）、パケット開始コードでない場合は、ステップ501に戻る。

【0066】また、パケット開始コードである場合は、パケット開始コードに含まれているデータ種別に基づいて、パケット410内のデータ412のデータ種別を判定し（ステップ504）、データ412のデータ種別が動画データである場合は、動画データ412をファイル202に格納する（ステップ505）。

【0067】また、データ412のデータ種別が動画データでない場合は、パケット開始コードに含まれているデータ種別に基づいて、パケット410内のデータ412のデータ種別を判定し（ステップ506）、データ412のデータ種別が音声データでない場合は、ステップ501に戻る。

【0068】また、データ412のデータ種別が音声データである場合は、音声データ412をファイル203に格納する（ステップ507）。

【0069】このように、ファイル201に格納されているオリジナルのMPEG1形式の音声付き動画データが、動画データと音声データとに分離され、各々、ファイル202およびファイル203に格納されることとなる。

【0070】次に、削減処理302について、図6および図7を用いて説明する。

【0071】図6はMPEG1形式の動画データの構造図である。

【0072】図6に示すように、MPEG1形式の動画データは、シーケンス601、GOP（Group Of Picture）602、ピクチャ603、スライス604、MB（Macro Block）605、ブロック606の6層からなる階層構造となっている。

【0073】シーケンス601は、画素数やフレームレート等の一連の同じ属性を持つ画面グループであり、1個以上のGOP602からなる。

【0074】GOP602は、復号処理単位となる画面グループの最小単位であり、1個以上のピクチャ（フレーム）603からなる。

【0075】ピクチャ603には、1枚の画面に共通な属性であり、Iピクチャ（Intra-Picture：フレーム内符号化画像）、Pピクチャ（Predictive-Picture：フレーム間順方向予測符号化画像）、Bピクチャ（Bidirectionally predictive-Picture：双方向予測符号化画像）の3個のピクチャタイプが存在する。なお、ピクチャ603は、1個以上のスライス604からなる。

【0076】ここで、各ピクチャタイプを構成するデータについて簡単に説明する。

【0077】Iピクチャを構成するデータは、フレーム間予測を使わずに、その情報だけから符号化されるデータであり、Pピクチャを構成するデータは、前方で1番

近い位置にあるIピクチャまたはPピクチャを構成するデータからの予測を行うことによってできるデータである。

【0078】また、Bピクチャを構成するデータは、前後の両方向で1番近い位置にあるIピクチャまたはPピクチャを構成するデータからの予測を行うことによってできるデータである。そこで、Bピクチャを構成するデータは、IピクチャおよびPピクチャを構成するデータが符号化された後に符号化されると共に、他のピクチャを符号化する際の予測に用いられない。なお、IピクチャまたはPピクチャを構成するデータは、周期的に出現するようになっている。

【0079】このように、MPEG1形式の動画データにおいては、Bピクチャを構成するデータが符号化される順序が異なることから、復号する順序と再生する順序とが異なることとなる。このため、正しい順序で復号および再生を行い、対応する音声と同時に出力することを可能とするために、上述したタイムスタンプが設けられている。

【0080】図6に戻って、スライス604は、1枚の画面を任意の長さで分割した小画面に共通の情報であり、画面の走査順に連続する1個以上のMB605からなり、MB605は、最大6個のブロック606からなる。

【0081】ブロック606は、2個以上の可変長符号607からなる。なお、可変長符号607は、空間周波数を示す係数であり、最終の可変長符号607は、EOB (End Of Block) 符号である。

【0082】図7は削減処理302の処理手順を示すフローチャートである。

【0083】削減処理302においては、図7に示すように、まず、所定の要求符号量に基づいて、ファイル204に格納すべき動画データの、ブロック606当りの可変長符号507の最大個数（以下、最大符号数と称す。）を決定する（ステップ701）。

【0084】ここで、要求符号量は、CPU101が、音声付き動画データ作成処理を行うに当たって、MPEG1形式の音声付き動画データを伝送すべきクライアントの処理能力やネットワークの伝送速度に基づいて決定しているものである。すなわち、CPU101は、クライアントの処理能力やネットワークの伝送速度に基づいて、対応する動画と音声とをクライアントで同時に出力することが可能な動画データの最大符号量を算出し、算出した最大符号量を要求符号量として決定する。なお、ビデオサーバがキーボードやマウス等の入力装置を備えている場合は、入力装置から入力されたものであってもよい。

【0085】続いて、ファイル202に格納されている動画データを、ブロック606を検出するまで走査し、ブロック606以外の部分を抽出してファイル204に

格納する（ステップ702）。

【0086】続いて、ステップ702で検出したブロック606に含まれている可変長符号607の個数をカウントし、符号数を取得する（ステップ703）。

【0087】続いて、ステップ701で決定した最大符号数とステップ703で取得した符号数とを比較し（ステップ704）、符号数が最大符号数よりも大きい場合は、ステップ702で検出したブロック606の先頭から最大符号数番目までの可変長符号607とEOB符号とをファイル204に格納する（ステップ705）。

【0088】また、符号数が最大符号数以下である場合は、ステップ702で検出したブロック606に含まれている全ての可変長符号607（EOB符号を含む。）をファイル204に格納する（ステップ706）。

【0089】最後に、ステップ702で検出したブロック606の次に、MPEG1ビデオ符号化規格で定められた動画データの終了を示す終了コードがあるか否かを判定し（ステップ707）、終了コードがある場合は、処理を終了し、終了コードがない場合は、ステップ702に戻る。

【0090】このように、ファイル202に格納されている動画データの一部（ブロック606内の可変長符号607）が削除されることによってデータ量が削減された動画データが、ファイル204に格納されることとなる。

【0091】なお、削減処理302によれば、EOB符号に近い可変長符号607が削除されるようになっているので、結果的には、直交変換方式で符号化された動画データ中の高周波数成分が削除されることとなる。

【0092】次に、多重処理303について、図8～図12を用いて説明する。

【0093】図10はMPEG1形式の音声データの構造図である。

【0094】図10に示すように、MPEG1形式の音声データは、複数のAAU (AudioAccess Unit) 1001からなる。

【0095】AAU1001は、復号処理の最小単位である所定のサンプル数の音声データを含み、AAUヘッダ1002とオーディオデータ1003とからなる。

【0096】ここで、AAUヘッダ1002は、符号化モード、ビットレート、サンプリング周波数等からなり、オーディオデータ1003は、符号化された音声データからなる。

【0097】図12は多重処理303の処理手順を示すフローチャートである。

【0098】多重処理303においては、図12に示すように、まず、ファイル204に格納されている動画データを解析し、GOP602ごとに、図8に示すGOP補助データ800を作成し、ピクチャ603ごとに、図9に示すピクチャ補助データ900を作成する（ステッ

ブ1201)。

【0099】図8に示すように、GOP補助データ800は、GOP602の再生開始時刻803と、GOP602の開始アドレス801と、GOP602の終了アドレス802とからなる。

【0100】GOP補助データ800において、再生開始時刻803は、動画データのフレームレートおよび前にある全てのGOP602に含まれているピクチャ603の数に基づいて算出することができる。また、開始アドレス801は、GOP602の先頭の、動画データの先頭からの位置に基づいて算出することができる。また、終了アドレス802は、次のGOP602の先頭の、動画データの先頭からの位置に基づいて算出することができる。なお、GOP602の次にGOP602がない場合は、終了アドレス802は、動画データの終端の、動画データの先頭からの位置に基づいて算出することができる。

【0101】また、図9に示すように、ピクチャ補助データ900は、ピクチャ603の再生開始時刻903と、ピクチャ603の復号開始時刻904と、ピクチャ603のピクチャタイプ902と、ピクチャ603の開始アドレス901とからなる。

【0102】ピクチャ補助データ900において、開始アドレス901は、ピクチャ603の先頭の、動画データの先頭からの位置に基づいて算出することができる。また、ピクチャタイプ902には、ピクチャ603のピクチャタイプをそのまま設定する。また、再生開始時刻903は、動画データのフレームレート、GOP補助データ800の再生開始時刻803、GOP602内でのピクチャ603の表示順序に基づいて算出することができる。また、復号開始時刻904は、動画データのフレームレート、GOP補助データ800の再生開始時刻803、GOP602内でのピクチャ603の表示順序、ピクチャタイプ902に基づいて算出することができる。

【0103】続いて、ファイル203に格納されている音声データを解析し、図11に示すAAU補助データ1100を作成する(ステップ1202)。

【0104】図11に示すように、AAU補助データ1100は、AAU1001の再生開始時刻1101と、AAU1001の開始アドレス1102とからなる。

【0105】AAU補助データ1100において、再生開始時刻1101は、音声データのサンプリング周波数およびAAU1001当りのサンプル数に基づいて算出することができる。また、開始アドレス1102は、AAU1001の先頭の、音声データの先頭からの位置に基づいて算出することができる。

【0106】続いて、対応する動画と音声とを同時に出力することを可能とするために、GOP602ごとに、該GOP602にAAU1101を対応付ける(ステッ

ブ1203)。

【0107】これは、各GOP602に対応するGOP補助データ800の再生開始時刻803に等しいかまたは1番近い再生開始時刻1101が設定されているAAU補助データ1100を検索していき、検索したAAU補助データ1100に対応するAAU1101から次に検索したAAU補助データ1100に対応するAAU1101までを、1個のGOP602に対応付ければよい。通常、1個のGOP602には、数十個のAAU1101が対応付けられることとなる。

【0108】続いて、多重対象として抽出すべきGOP602、および、ステップ1203で該GOP602に対応付けられたAAU1001を決定する(ステップ1204)。

【0109】ここで、多重対象として抽出すべきGOP602とは、MPEG1形式の音声付き動画データを伝送すべきクライアントから指定された再生速度に基づいて決定されるものである。すなわち、例えば、クライアントから指定された再生速度が1倍速(通常の再生速度)である場合は、全てのGOP602を多重対象として抽出すればよい。また、例えば、クライアントから指定された再生速度が2倍速(早送り再生の再生速度)である場合は、1/2の再生時間で再生することができるようにするために、全てのGOP602から、GOP602を1個ずつスキップするようにして、半分のGOP602を多重対象として抽出すればよい。

【0110】一般的には、S個のGOP602からT個のGOP602を多重対象として抽出し、抽出したGOP602の各々について、対応付けられたAAU1001を決定することにより、オリジナルの再生時間のT/S倍の再生時間分の音声付き動画データを作成することが可能である。ここで、SおよびTは、共に自然数であり、 $S \leq T$ である。

【0111】最後に、ステップ1204で決定したGOP602およびAAU1001を、各々、パケットに分割してから多重することにより、MPEG1形式の音声付き動画データを作成し、作成したMPEG1形式の音声付き動画データをファイル205に格納する(ステップ1205)。

【0112】なお、このとき、バックヘッダ401に含まれるシステム時刻基準参照値、および、パケットヘッダ411に含まれるタイムスタンプを、ピクチャ補助データ900の再生開始時刻903およびAAU補助データ1100の再生開始時刻1101から算出して設定する。

【0113】また、パケットに分割する際に、1個のパケット410に格納するデータ412の種別を、ピクチャ補助データ900の再生開始時刻903および開始アドレス901、並びに、AAU補助データ1100の再生開始時刻1101および開始アドレス1102に基づ

10

20

30

40

50

いて決定するようにする。

【0114】このように、ファイル204に格納されている動画データとファイル203に格納されている音声データとが多重された、所望の再生時間分のMPEG1形式の音声付き動画データがファイル205に格納されることとなる。

【0115】以上説明したように、本実施例によれば、MPEG1形式のオリジナルの音声付き動画データから、該オリジナルの音声付き動画データよりデータ量が少なく、かつ、対応する動画と音声とを同時に出力するのに適した、所望の再生時間分のMPEG1形式の音声付き動画データを作成することができる。

【0116】なお、本実施例では、作成したMPEG1形式の音声付き動画データを、一旦記憶装置103に格納してからクライアントに伝送するようにしているが、作成したMPEG1形式の音声付き動画データを、そのままクライアントに伝送するようにしてもよい。

【0117】次に、本発明の第2の実施例について説明する。

【0118】第2の実施例は、上記第1の実施例における削減処理302の代わりに、以下に説明する削減処理304を行うようにしたものである。

【0119】以下、削減処理304について、図13および図14を用いて説明する。

【0120】図14は削減処理304の処理手順を示すフローチャートである。

【0121】削減処理304においては、図14に示すように、まず、ファイル202に格納されている動画データを、ピクチャ603を検出するまで走査し、ピクチャ603以外の部分を抽出してファイル204に格納する(ステップ1401)。

【0122】続いて、ステップ1401で検出したピクチャ603のピクチャタイプに基づいて、該ピクチャ603がBピクチャであるか否かを判定し(ステップ1402)、Bピクチャである場合は、Bピクチャを構成するデータの代わりに、図13に示すダミーデータをファイル204に格納する(ステップ1403)。

【0123】図13はダミーデータの構造を示す図である。

【0124】図13において、ピクチャ開始コード1301は、ピクチャ603の開始を示すピクチャ開始コード(同期コードの一種ある。)である。また、表示順序1302には、ピクチャ603のGOP602における表示順序が設定される。また、ピクチャタイプ1303は、ピクチャ603のピクチャタイプを示すデータであり、ここでは、ピクチャタイプ1303には、Bピクチャであることを示すデータが設定される。

【0125】また、IピクチャまたはPピクチャである場合は、IピクチャまたはPピクチャを構成するデータをそのままファイル204に格納する(ステップ140

4)。

【0126】最後に、ステップ1401で検出したピクチャ603の次に動画データの終了を示す終了コードがあるか否かを判定し(ステップ1405)、終了コードがある場合は、処理を終了し、終了コードがない場合は、ステップ1401に戻る。

【0127】このように、ファイル202に格納されている動画データの一部(Bピクチャを構成するデータ)をデータ量が少ない所定のデータ(ダミーデータ)に置き換えることによってデータ量が削減された動画データが、ファイル204に格納されることとなる。

【0128】なお、削減処理304において、Bピクチャを構成するデータのみをダミーデータに置き換えているのは、上述したように、Bピクチャを構成するデータは、他のピクチャを符号化する際の予測に用いられないので、Bピクチャを構成するデータを削除しても、他のピクチャの画質に影響を与えないからである。

【0129】以上説明したように、本実施例によれば、MPEG1形式のオリジナルの音声付き動画データから、該オリジナルの音声付き動画データよりデータ量が少なく、かつ、対応する動画と音声とを同時に出力するのに適した、所望の再生時間分のMPEG1形式の音声付き動画データを作成することができる。

【0130】

【発明の効果】以上説明したように、本発明によれば、音声付き動画データから、該音声付き動画データよりデータ量が少なく、かつ、対応する動画と音声とを同時に出力するのに適した、所望の再生時間分の音声付き動画データを作成することができるという効果がある。

【図面の簡単な説明】

【図1】第1の実施例を適用したビデオサーバの構成を示すブロック図。

【図2】第1の実施例の音声付き動画データ作成処理におけるデータの流れを示す説明図。

【図3】第1の実施例の音声付き動画データ作成処理の処理手順を示すフローチャート。

【図4】MPEG1形式の音声付き動画データの構造図。

【図5】第1の実施例における分離処理の処理手順を示すフローチャート。

【図6】MPEG1形式の動画データの構造図。

【図7】第1の実施例における削減処理の処理手順を示すフローチャート。

【図8】第1の実施例における削減処理で用いるGOP補助データの構造図。

【図9】第1の実施例における削減処理で用いるピクチャ補助データの構造図。

【図10】MPEG1形式の音声データの構造図。

【図11】第1の実施例における削減処理で用いるAAU補助データの構造図。

【図12】第1の実施例における多重処理の処理手順を示すフローチャート。

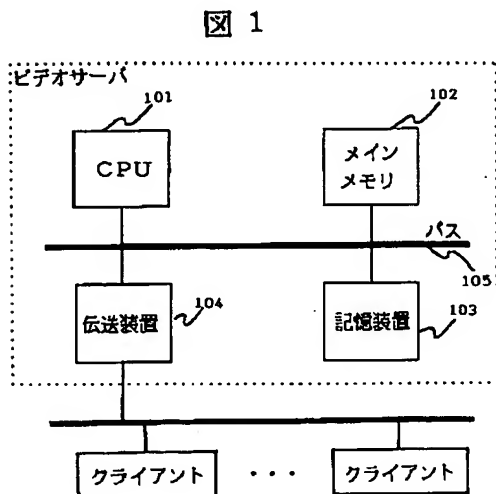
【図13】第2の実施例における削減処理で用いるダメージデータの構造図。

【図14】第2の実施例における削減処理の処理手順を示すフローチャート。

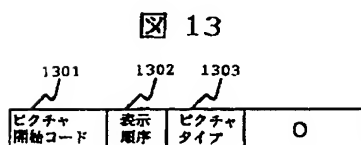
【符号の説明】

101…CPU、102…メインメモリ、103…記憶装置、104…伝送装置、105…バス、301…分離処理、302、304…削減処理、303…多重処理。

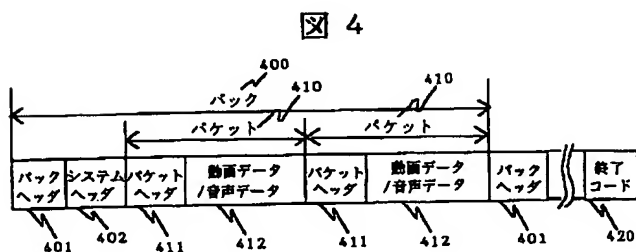
【図1】



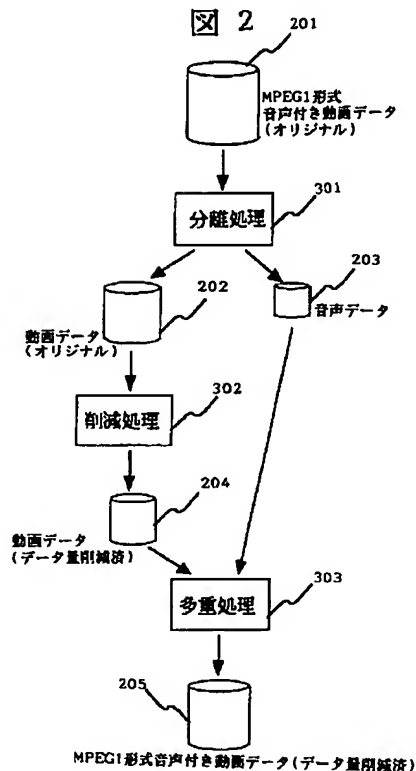
【図13】



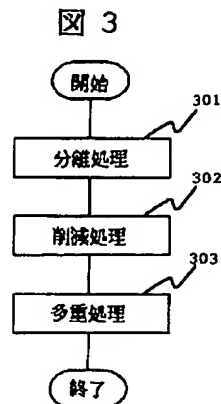
【図4】



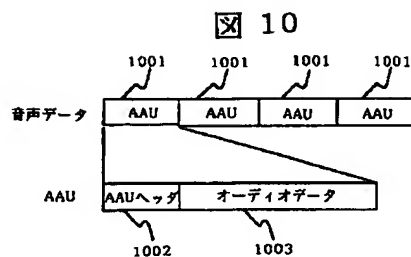
【図2】



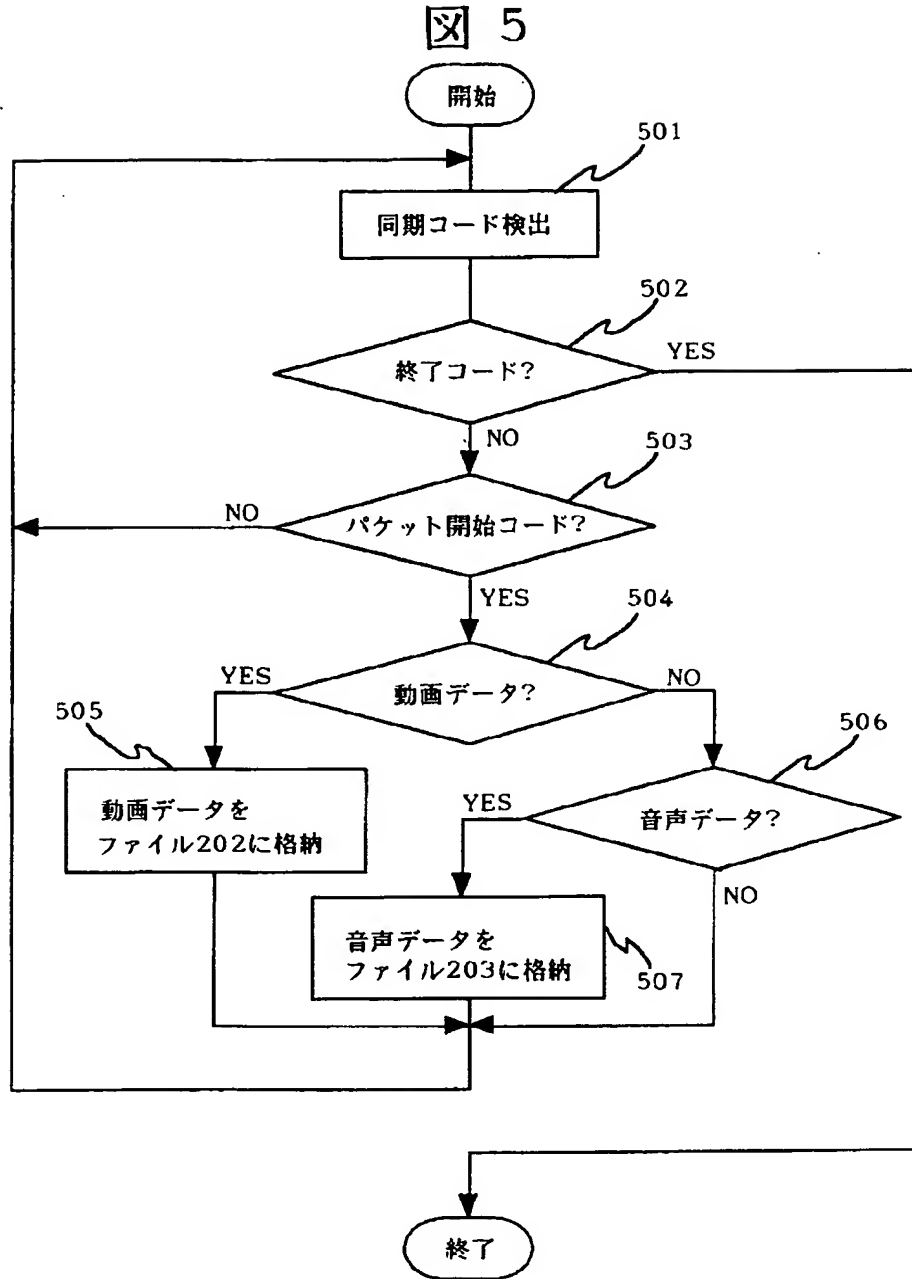
【図3】



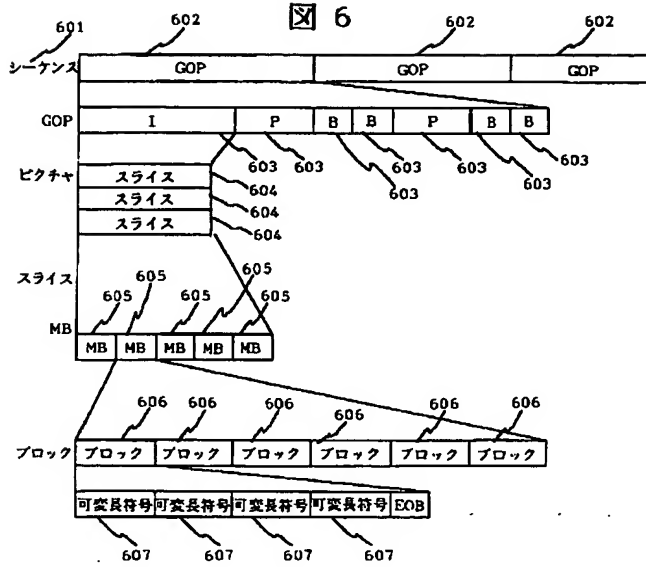
【図10】



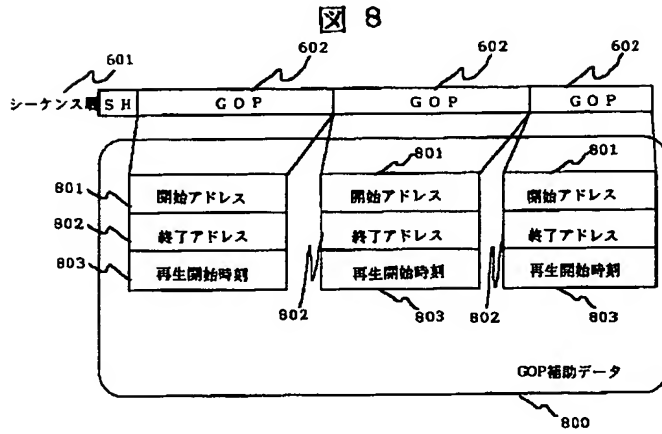
【図 5】



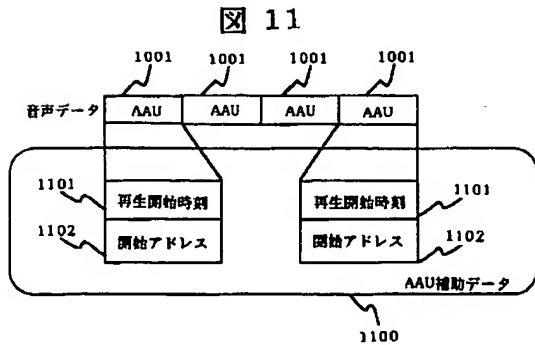
【図6】



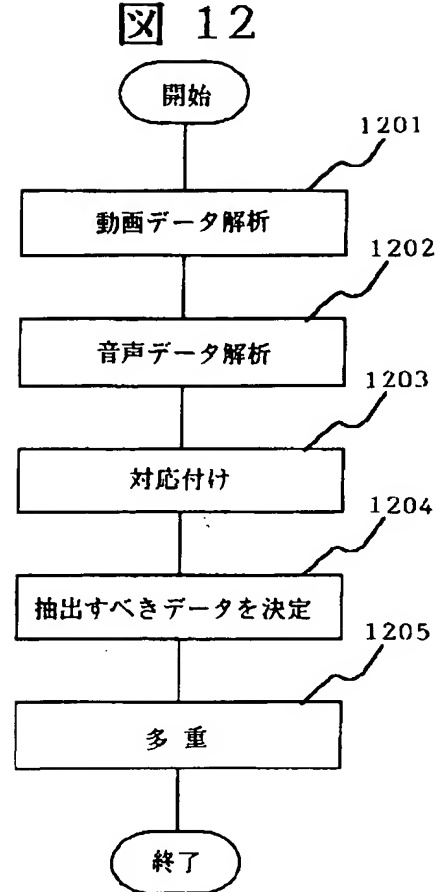
【図8】



【図11】

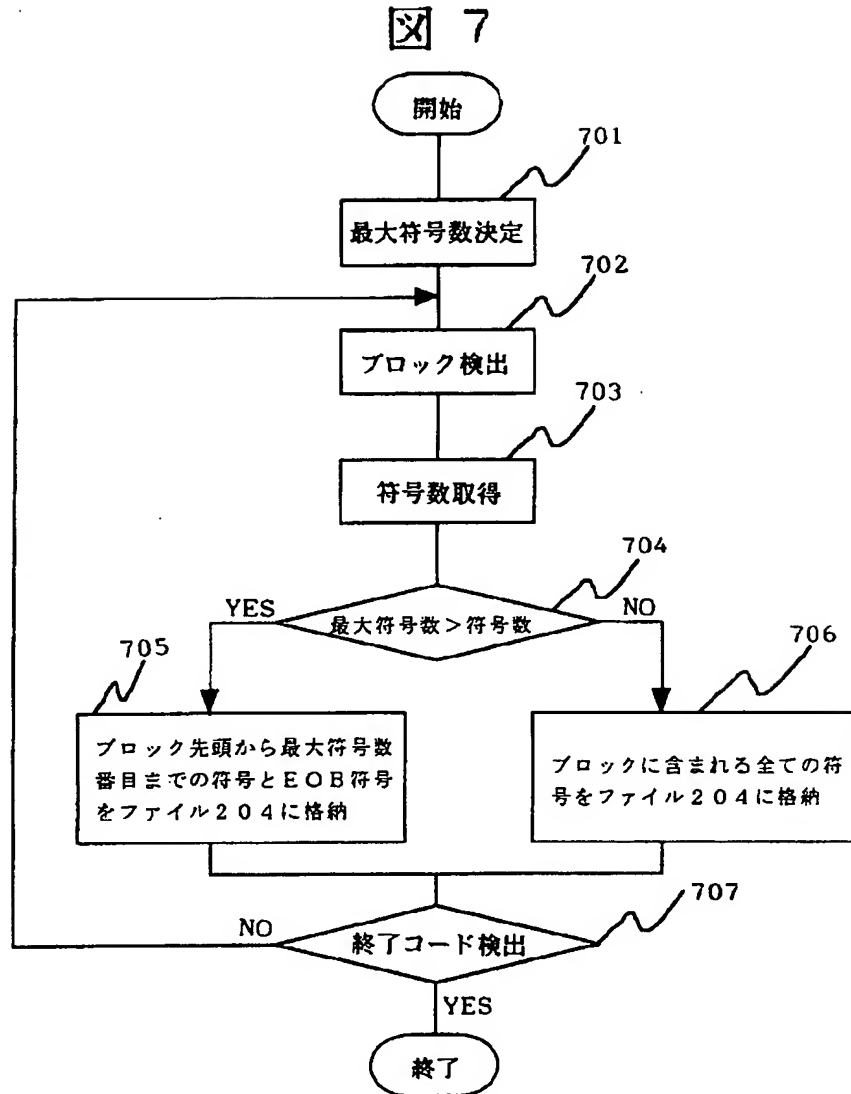


【図12】

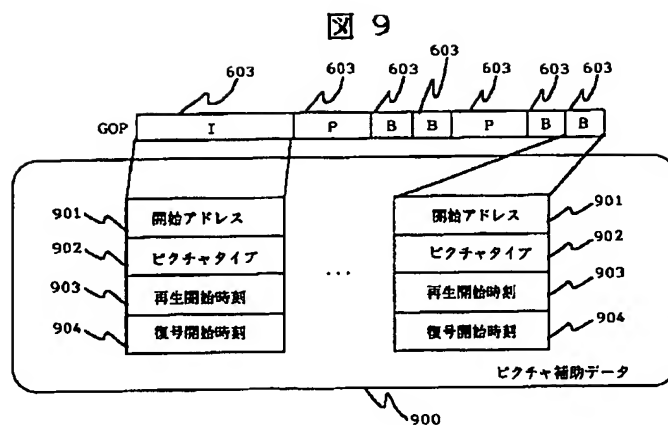




【図7】

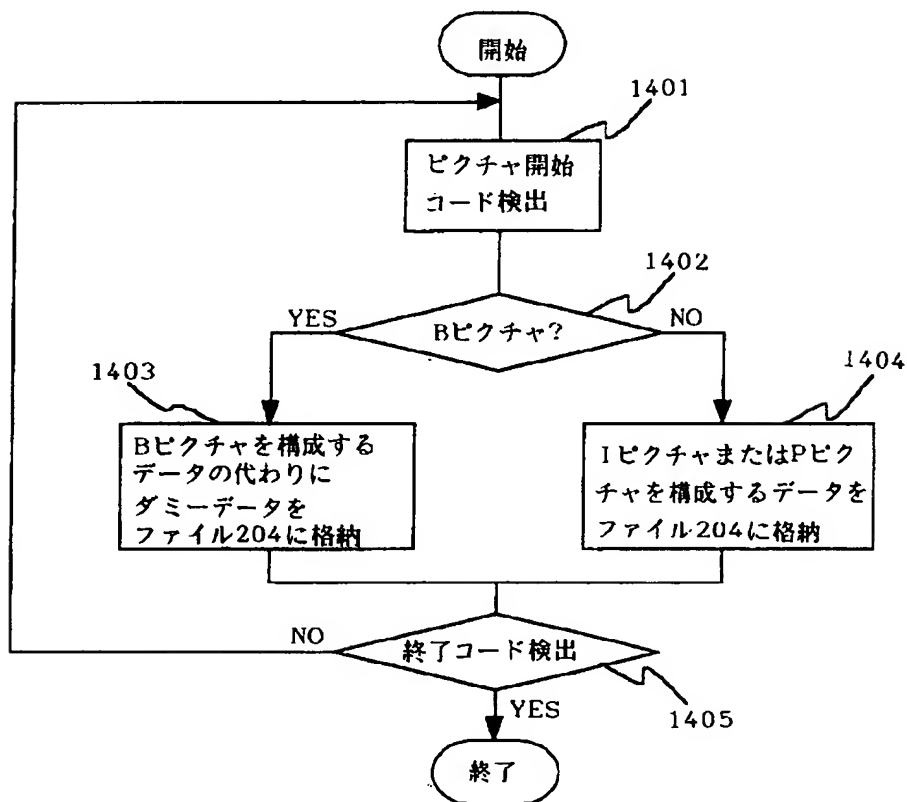


【図 9】



【図 14】

図 14



フロントページの続き

(72)発明者 山田 剛裕  
神奈川県川崎市麻生区王禅寺1099番地 株  
式会社日立製作所システム開発研究所内

(72)発明者 田中 和明  
神奈川県横浜市戸塚区戸塚町5030番地 株  
式会社日立製作所ソフトウェア開発本部内  
(72)発明者 今井 康裕  
神奈川県海老名市下今泉810番地 株式会  
社日立製作所オフィスシステム事業部内